La **inferencia** es la rama de la Estadística que tiene técnicas apropiadas para dar solución a estos tipos de problemas.

Existen sucesos llamados aleatorios porque son resultados de experimentos que dependen del azar.
Ejemplos:

- Lanzamiento de una moneda o de un dado
- Sexo de un recién nacido
- Medir presión arterial diastólica

Las **variables aleatorias** son los resultados de sucesos aleatorios
Son variables de las que conocemos puede tener un resultado dentro de un intervalo de valores posibles, pero no conocemos con certeza cuál será ese resultado, aún y cuando se repita el experimento bajo las mismas condiciones.

Sexo tiene sólo dos valores posibles mientras presión arterial diastólica puede tomar cualquier valor dentro de un rango de valores posibles.
A las variables del tipo como el sexo, que sólo toman algunos valores se les llama **discretas** mientras a las del tipo de la presión arterial diastólica se les llama **continuas.**

PROBABILIDAD:
Vamos a continuar trabajando con la variable aleatoria "sexo", que tiene como resultados: F y M. podemos decir que el suceso A = (obtener M) es el que nos interesa. Si en un estudio de tamaño n se observan k repeticiones del suceso A, entonces la probabilidad del suceso (P(A)) se define como la frecuencia relativa de A.

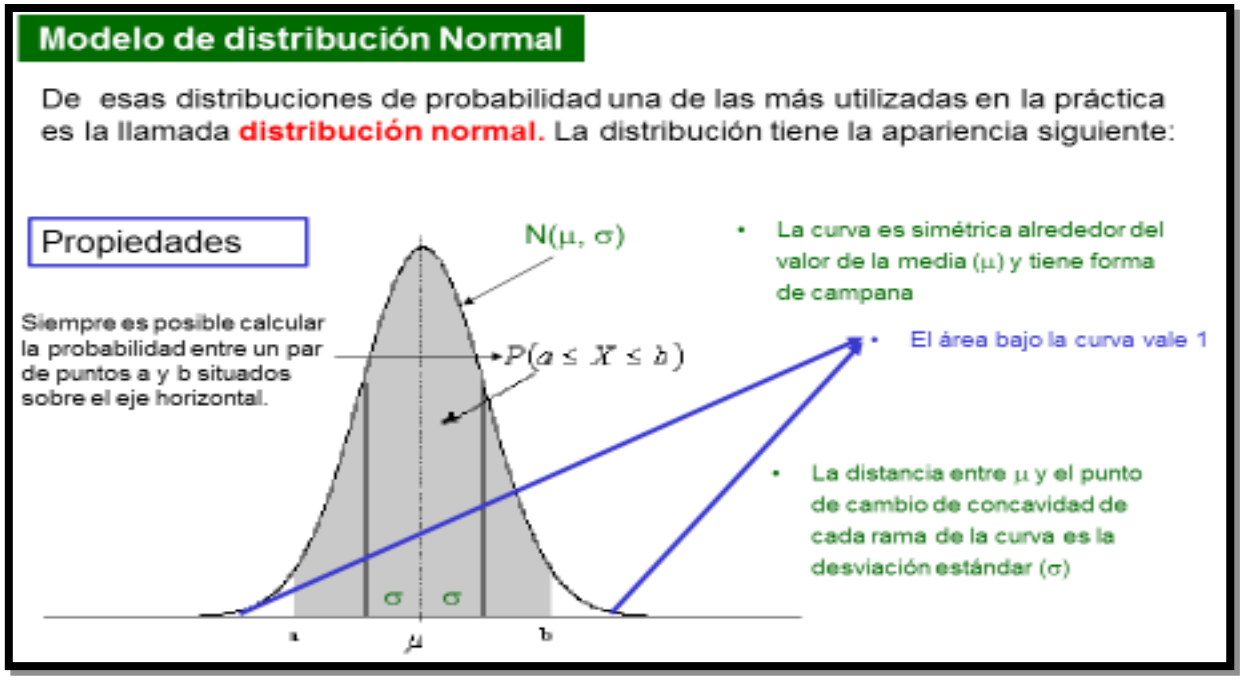
- $P(A) = k/n$

En la probabilidad se cumple que:

- $0 \leq P(A) \leq 1$
- $P(Ac) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Si $A \cap B$ = vacía, entonces $P(A \cup B) = P(A) + P(B)$.
- Si B es subconjunto de A, entonces $P(B) \leq P(A)$.
- La suma de las probabilidades de todos los sucesos elementales es 1 (sucesos elementales en este caso sexo masculino y sexo femenino)

El **modelo teórico de distribución o ley de la variable aleatoria** es la curva a la que tiende el polígono de frecuencias relativas a medida que se acumula un número mayor de valores de la variable aleatoria, en este caso peso del niño al nacer.

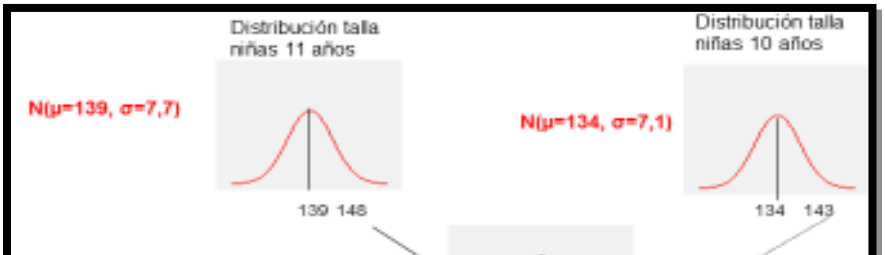
Modelo de distribución Normal
Para cada par de valores μ y σ se puede graficar una curva, es decir que la forma de la distribución queda definida a partir de esos valores**. Estos valores son los parámetros de la distribución** y se corresponden con la media y la desviación estándar poblacional de la variable.


Modelo de distribución Normal
De esas distribuciones de probabilidad una de las más utilizadas en la práctica es la llamada distribución normal. La distribución tiene la apariencia siguiente:
Propiedades
Siempre es posible calcular la probabilidad entre un par de puntos a y b situados sobre el eje horizontal.
N(μ, σ)
P(a ≤ X ≤ b)
σ
σ
a
μ
b
La curva es simétrica alrededor del valor de la media (μ) y tiene forma de campana
El área bajo la curva vale 1
La distancia entre μ y el punto de cambio de concavidad de cada rama de la curva es la desviación estándar (σ)

Modelo de distribución Normal estándar
Problema:
Supongamos que tenemos dos niñas, una de ***10 años con 143 cms de talla que proviene de una población con promedio de talla de 134 cms y desviación estándar de 7,1 cms***; y la otra niña tiene ***11 años con una talla de 148 cms que proviene de una población con promedio de talla de 139 cms y desviación estándar de 7,7 cms***. ¿cuál de las dos niñas es más alta en relación a la población de procedencia (está más alejada del promedio de su grupo)?
Solución:
$Z_{niña10años}$ = (talla niña 10 – media) /desviación estándar= (143 – 134) / 7,1 = 9 / 7,1 = 1,27
$Z_{niña11años}$ = (talla niña 11- media) /desviación estándar= (148 – 139) / 7,7 = 9 / 7,7 = 1,17
R: La niña más alta es la de 10 años

Para poder hacer comparaciones entre características de individuos o grupos de individuos que provengan de distribuciones normales diferentes y para facilitar el cálculo de los valores de área bajo la curva se usa la **estandarización** de la curva normal a una **curva con media cero y desviación estándar uno.**

Entonces en el problema anterior 1.27 es el valor

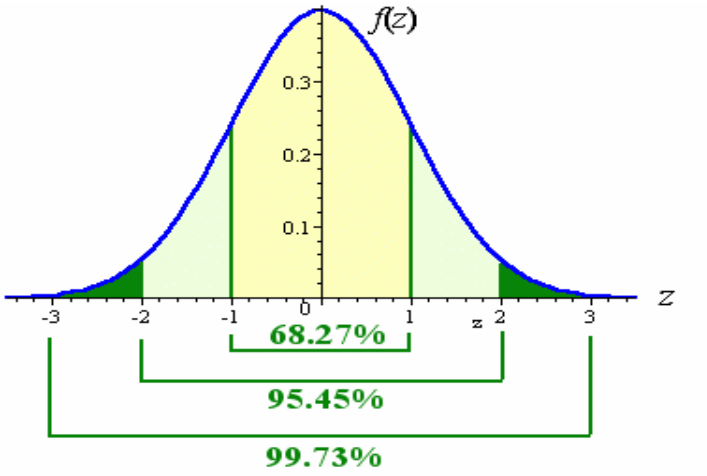


estandarizado correspondiente a la talla de la niña de 10 años y 1.17 el correspondiente a la talla de la niña de 11 años.

Por convención se usa la letra z para referirnos a la distribución normal estándar donde las puntuaciones estándar serían:

Z = (x-m) /s

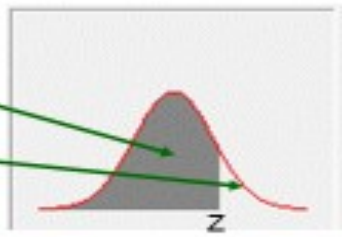
Existen tablas donde ya aparece calculada el área bajo la curva para diferentes valores de z.



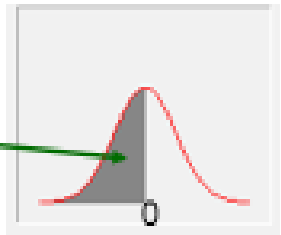
Revisemos las propiedades de la probabilidad suponiendo que el suceso estudiado sigue una distribución normal y usando la normal estandarizada (media =0 y desviación estándar=1).

El área bajo la curva es 1.

El área oscura bajo la curva es **P(Z<z)** y el área clara su complemento es decir **P(Z>=z) = 1 - P(Z<z)**

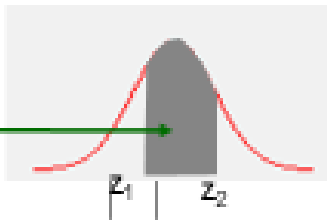


El área bajo la curva hasta z=0 es 0.5.



Para dos valores $z_1 < z_2$ se cumple que

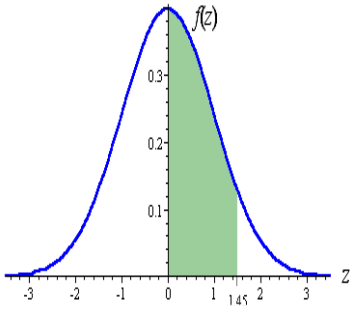
$P(z_1 < Z < z_2) = P(Z < z_2) - P(Z < z_1)$



ESTADÍSTICA INFERENCIAL
RANDY MEJÍAS GONZÁLEZ

Modelo de distribución Normal tipificada. Usos de la tabla de áreas bajo la curva

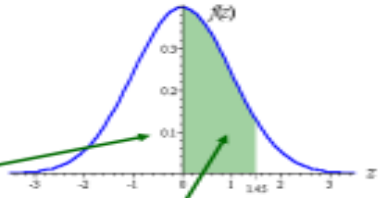
Ejemplo 1:



Supongamos que interesa conocer que proporción de valores caen entre dos puntajes z. El área verde (en este tipo de problemas es necesario dibujar a mano un gráfico que nos ayude a visualizar lo que estamos haciendo) está delimitada por los puntajes 0 y 1.45. Para saber cuál es el valor de esta área tenemos que referimos a la tabla de la distribución normal estándar.

Caso 1: Calcular la probabilidad α (área sombreada)

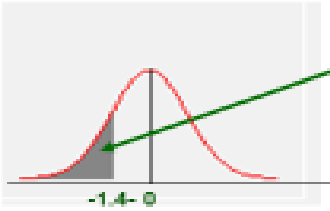
P (Z< z=1.45) = 0.928



El área bajo la curva hasta z=0 es 0.5.

P (0<Z< 1.45) = P (Z<1.45) - P (Z< 0) = 0.928-0.5 = 0.428

Caso 2:



P (Z< -1.4) = 0.081

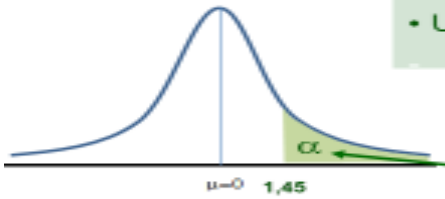
Calcular la probabilidad a la izquierda de un valor negativo de z.

•Usamos la tabla de la distribución

Caso 3:

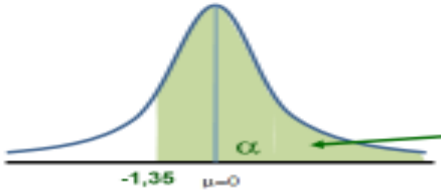
Calcular el área a la derecha de un valor Z

- Aplicamos las propiedades de las probabilidades
- Usamos la tabla de la distribución



P (Z>1,45))=1-P (Z< 1,45) = 1- 0,928

= 0,072

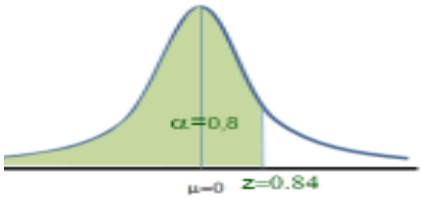


P (Z>-1,35))= 1 - P(Z<-1.35) = 1 - 0.9 = 0,1

Caso 4: Conocer el valor de z por debajo del cuál se encuentra el área

Ejemplo:

Ahora lo que interesa **es conocer el valor de z por debajo del cuál se encuentra el 80% del área**. Es decir queremos encontrar $Z_{0.8}$



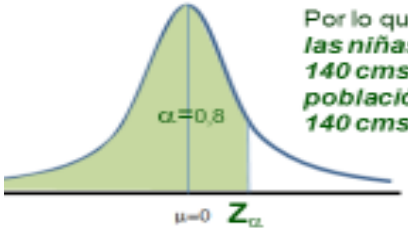
Tendremos que buscar en las columnas Área un valor igual o aproximado a 0.8.

El valor z = 0.84 cumple con esa caracteristica.

Si la variable tipificada corresponde a una población de tallas de niñas de 10 años **con promedio 134 cms y desviación típica de 7 cms.**

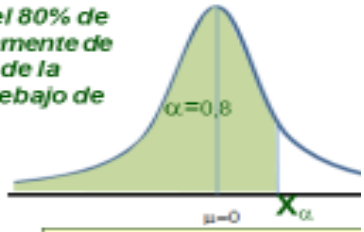
Despejando $Z = (x-\mu)/\sigma$

$X = 134 + 0.84*7; \quad X = 134 + 5.88 = 139.88$



$Z_{\alpha} = 0,84$

Por lo que podemos decir que ***en esa población el 80% de las niñas tendrán tallas por debajo aproximadamente de 140 cms o que la probabilidad de que una niña de la población extraída al azar tenga una talla por debajo de 140 cms es 0.8.***



Entonces $X_{\alpha} = 139,88$

GENERALIZANDO:

Para hallar el puntaje Z, es decir calcular a partir de x su correspondiente en **Z~N(0,1)**
utilizamos la fórmula: $Z = (x-m)/s$

Para calcular el punto x correspondiente a Z en X**~N(**m,s**)** utilizamos entonces:
$X = m + Zs$

DISTRIBUCIÓN T-STUDENT

Otra función de probabilidad conocida, también para variables continuas es la distribución **t de Student.**

- Tiene un parámetro denominado grados de libertad que se puede asociar al tamaño de la muestra .
- Es algo más dispersa que la normal.
- Mientras más aumenta el tamaño de la muestra más se acerca a N(0,1).
- Es simétrica con respecto al cero, que es su valor medio.

Población

Conjunto finito de objetos de cierto tipo; seres vivos, viviendas, policlínicos, hospitales, por poner algunos ejemplos. Es el grupo sobre el cual el investigador está interesado en ofrecer conclusiones.

Muestra

Es un subconjunto de la población. Generalmente se usa cuando la población es demasiado grande para poder estudiarla en su totalidad o cuando es imposible hacerlo.

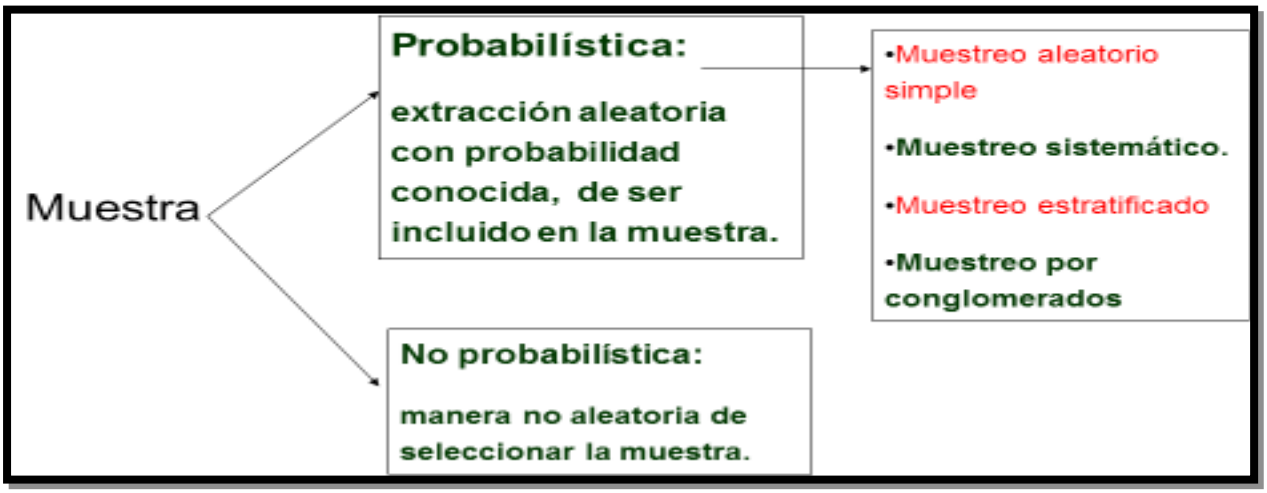
Ventajas del muestreo:

- Disminución de los costos de la investigación.
- Reducción del tiempo de procesamiento y análisis de los resultados.
- Ampliación del campo de posibilidades de realizar determinados estudios.
- Permite usar para la recogida de información instrumentos mejor diseñados y un personal mejor capacitado.

La **desventaja fundamental del muestreo** es el llamado **error del muestreo**

- Una de las causas: ausencia de "Representatividad
- Otra causa que se puede controlar, pero no evitar: variabilidad inherente al fenómeno que se estudia.

Población y muestra. Conceptos de estimador y distribución muestral



Parámetros poblacionales.
Valores que caracterizan la ley de distribución de una variable aleatoria. Para la ley normal los parámetros son la media y la desviación estándar. Para denominarlos se usan letras griegas, para la media m y para las desviación estándar s.
Estimador.
Expresión matemática aplicada a los valores de una **muestra aleatoria de tamaño n** con el objetivo de estimar un parámetro poblacional.
Ejemplos: media aritmética $\bar{x}=\sum x_i/n$

la desviación estándar $s=\sqrt{\sum (x_i-\bar{x})^2/n}$

Los **estimadores son también variables aleatorias** porque dependen de muestras aleatorias.
Las medias de todas las muestras posibles de tamaño n formarán una nueva población de variables aleatorias, en este caso la población de las medias.
Ahora se puede considerar la distribución de probabilidades de la variable "media". A esa nueva distribución de probabilidades es a lo que se denomina **Distribución muestral de la media.**
A la **desviación estándar de la media muestral** se le conoce por el nombre de **error estándar de la media.**

Relación entre población y muestra.

Para una población cualquiera de tamaño N con media m y desviación estándar s, la **distribución muestral de la media**, calculada a partir de todas las muestras aleatorias de **tamaño n**, será normal con media m y desviación estándar $\sigma\sqrt{n}$. La desviación estándar de la media. muestral recibe el nombre **de error estándar**.

Dicho con otras palabras si X variable aleatoria con distribución de parámetros m y s; la media muestral para muestras de tamaño n, tendrá distribución con la misma media que X y un error estándar menor que la desviación estándar, pues es la desviación de la variable aleatoria original dividida por la raíz cuadrada de n.

==========

Los parámetros poblacionales son valores que caracterizan la ley de distribución de una variable aleatoria.
Para la ley normal los parámetros son la media (μ) y la desviación estándar (σ).

¿Qué es estimación?

Supongamos que queremos realizar un estudio de una población cualquiera de la que desconocemos el valor de alguno de sus parámetros, por ejemplo, su media poblacional. Podemos tomar una muestra aleatoria de dicha población para calcular una aproximación a este valor desconocido. A esa aproximación se llama estimación.

Como muy probablemente esa estimación no coincidirá con el valor real del parámetro, se debe tener en cuenta el error aproximado que se comete al realizarla

Tipos de estimaciones de parámetros:

Estimación puntual: Se busca un estimador, que con base a los datos muestrales de origen a un valor puntual que utilizamos como estimación del parámetro.
Estimadores por intervalos de confianza: Se determina un intervalo aleatorio que, de forma probable, contiene el verdadero valor del parámetro

Estimador o estadígrafo.
Expresión matemática aplicada a los valores de una **muestra aleatoria de tamaño n** con el objetivo de estimar un parámetro poblacional.
Ejemplos:

1. Estimador puntual de la media poblacional (μ): $\overline{x} = \sum x_i / n$

2. Estimador puntual de la desviación estándar poblacional (σ): $s = \sqrt{\sum (x_i - \overline{x})^2 / (n-1)}$

3. Estimador puntual de una proporción poblacional P: $\hat{p} = k/n$ *(donde k: número de elementos de la muestra con la característica de interés)*

Los **estimadores son también variables aleatorias** porque dependen de muestras aleatorias.

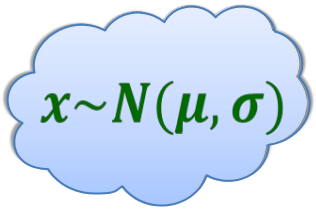
ESTADÍSTICA INFERENCIAL
RANDY MEJÍAS GONZÁLEZ

Las medias de todas las muestras posibles de tamaño n formarán una nueva población de variables aleatorias, en este caso la población de las medias.
Ahora se puede considerar la distribución de probabilidades de la variable "media". A esa nueva distribución de probabilidades es a lo que se denomina **Distribución muestral de la media.**
A la **desviación estándar de la media muestral** se le conoce por el nombre de **error estándar de la media.**

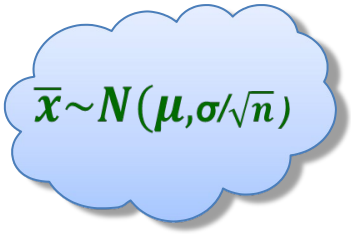
Relación entre población y muestra.

Para una población cualquiera de tamaño N con media m y desviación estándar σ, es decir, □~□(□, □):
la distribución muestral de la media, calculada a partir de todas las muestras aleatorias de tamaño n, será normal con media m y $\sigma/\sqrt{n}$ desviación estándar (error estándar) , o sea,

$$\bar{x} \sim N(\mu, \sigma\sqrt{\square}\ /\ )$$



Población de tamaño N donde la variable aleatoria de interés (x) tiene ley de distribución normal.



Muestra aleatoria simple única de tamaño n donde □ tiene ley de distribución normal.

Estimación puntual del parámetro desconocido

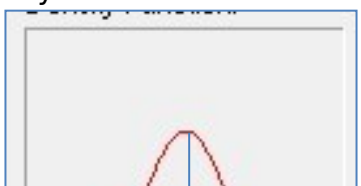
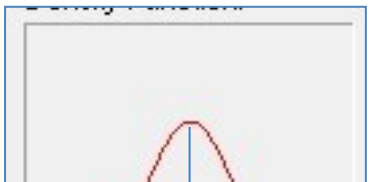
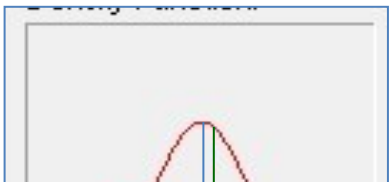
La estimación puntual del parámetro desconocido es un valor que se obtiene **evaluando** una función en los valores que tome la variable de interés **en una muestra** dada.

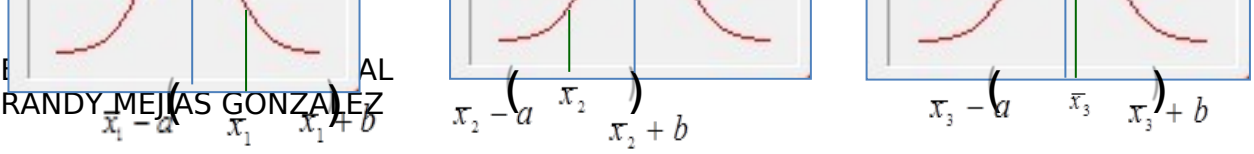
Note que dicha estimación es un valor único, que varía solo si se usa otra muestra, por eso se dice que es una estimación puntual del parámetro.
Como ejemplo, tenemos a $\bar{x}$ evaluada para un conjunto de datos

¿Creen ustedes que este estimador puntual sea la mejor solución al problema?

Cada vez que se obtenga una muestra nueva, aunque sea del mismo tamaño se obtendrá una estimación puntual diferente para el parámetro poblacional
Entonces: La estimación puntual no es la mejor solución al problema

Una solución:
En lugar de usar un solo valor introducir un intervalo de valores posibles para la estimación. Es decir, calcular un intervalo aleatorio que contenga al verdadero valor del parámetro desconocido con una cierta probabilidad. Para lograr intervalos de este tipo: a la media muestral, estimador de la media de la población, se le sustrae y se le añade valores a y b tal como se muestra a continuación:



¿Cómo escoger a y b?

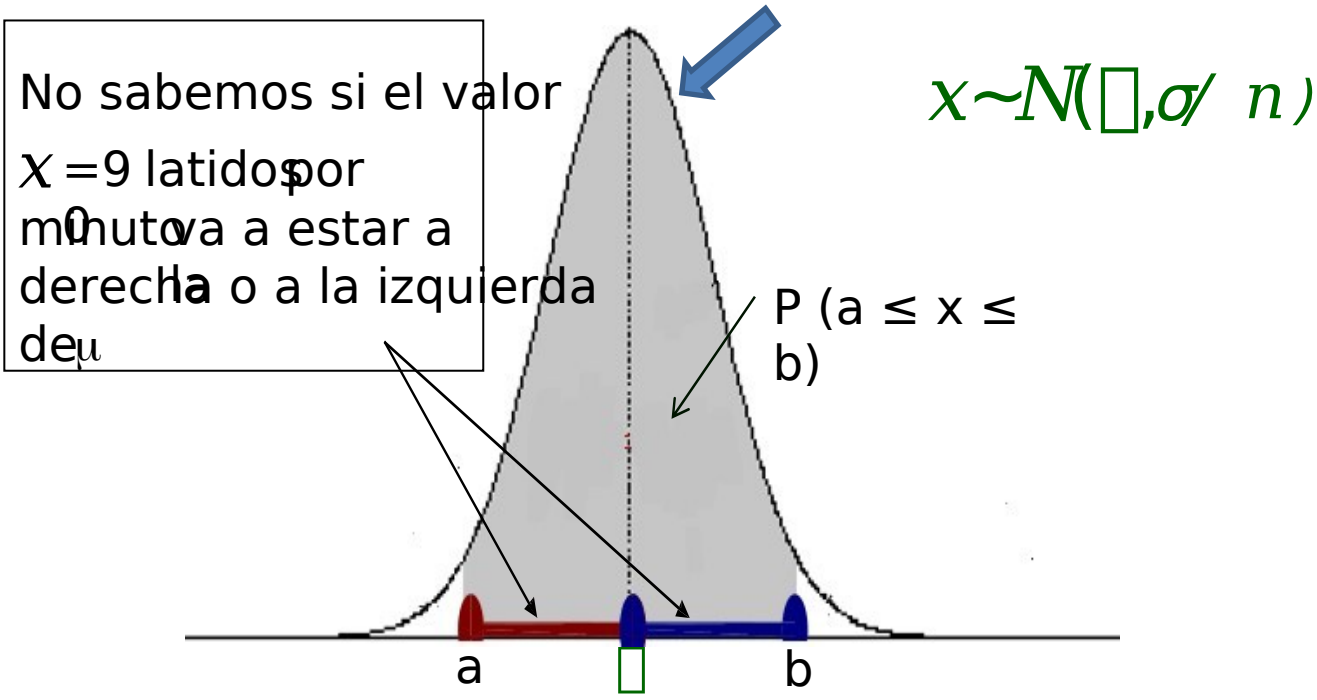
Población con distribución normal y varianza conocida

Problema a solucionar mediante un intervalo de confianza.

En un experimento diseñado para estimar el número medio de latidos del corazón por minuto de una cierta población, se obtuvo que el número medio de latidos por minuto para 49 sujetos fue de 90. Se supone que esos 49 sujetos son una muestra aleatoria simple de la población que a su vez tiene una distribución normal con una desviación estándar de 10 latidos por minuto. Halle una estimación por intervalo para la media poblacional.

CONOCEMOS QUE:

Distribución de muestras de tamaño n: $N(\mu, \sigma / \sqrt{n})$



Para la media. Población con distribución normal y varianza conocida

Nivel de Confianza (1- α) es probabilidad de que el verdadero valor del parámetro poblacional esté dentro del intervalo de confianza.

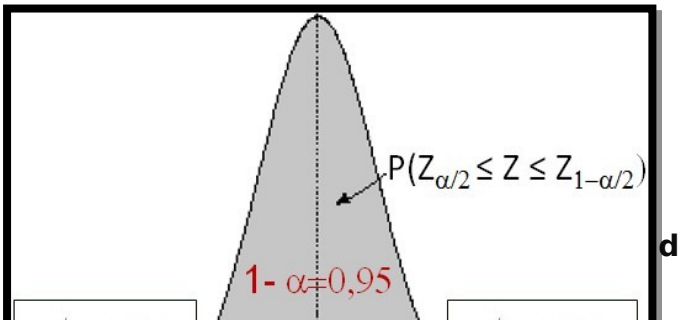
Nivel de significación (α) es la probabilidad de que el verdadero valor del parámetro poblacional no esté dentro del intervalo

EJEMPLO:

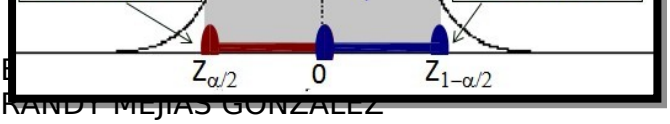
Suponga que queremos un Nivel de Confianza del 95%. Por tanto: 1-α= 0,95 y α =0,05

Distribución Normal standard

$Z \sim N$ **(0,1)**



Entonces α debe ser pequeño Lo más común es usar α = 0.05.



$Z \sim N(0,1)$

$$\bar{x} \sim N(\mu, \sigma/\sqrt{n})$$

Z=(□ - □/σ m)√

Partiendo de que Pr (Z0,025 < Z< Z0,975) = 0,95, sustituyendo Z y despejando para centrar el parámetro a estimar m se llega a la expresión del intervalo de confianza

$$\Pr\left\{\bar{x} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 0{,}95$$

Generalizando:

$$\Pr = \left\{\underbrace{\bar{x} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\text{Límite inferior=li}} < \mu < \underbrace{\bar{x} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\text{Límite superior=ls}}\right\} \; 1 - \alpha$$

Es decir que el **intervalo del (1-α)100% de confianza es de la forma (li; ls).** Se debe notar que mientras **la media poblacional es desconocida, pero de valor fijo, los límites del intervalo** son aleatorios y **cambiarán en función de la muestra que se tome**.

Ahora podremos aplicar la expresión hallada a los datos del ejemplo para calcular el IC del 95% para la media poblacional de latidos del corazón

Datos
n = 49;
desviación estándar de la población = 10
media de la muestra = 90
coeficiente de confianza z1-0,05/2 = z0,975=1.96

Sustituyendo en la fórmula del IC y buscando los percentiles en la tabla de la Distribución Normal Estándar

$$Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = (1.96)\left(\frac{10}{\sqrt{49}}\right) = 2.8$$

límite inferior= 90-2,8 = 87,2
límite superior= 90+2,8= 92,8

El intervalo obtenido es [87,2; 92,8]. Puede escribirse también 87,2≤μ≤92,8

Interpretación: Para un nivel de confianza del 95 % , la media poblacional de latidos del corazón fluctúa aproximadamente entre 87,2 y 92,8 latidos por minuto

Se ha construido un intervalo que contiene al verdadero valor del parámetro con una probabilidad de 0,95; si se toman 100 muestras de tamaño n fijo es de esperar que aproximadamente 95% de tales muestras contienen el verdadero valor del parámetro, mientras que alrededor del 5% no lo contienen
Recordemos que:
• El estimador por intervalo de confianza es un intervalo probabilístico que puede contener el verdadero valor del parámetro desconocido con determinada confianza.
• La estimación es cuando evaluamos el estimador en los valores de una muestra aleatoria dada.
Estimación

$$\left\{ \underbrace{\bar{x} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\text{Límite inferior}} < \mu < \underbrace{\bar{x} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\text{Límite superior}} \right\}$$

Resumiendo, El Intervalo del (1-α)100%de confianza (IC) para m con varianza conocida depende de:
1. La media $\bar{x}$ muestral
2. La desviación estándar poblacional σ
3. El tamaño de la muestra n
4. El valor que es el percentil de orden 1 – α/2 de la distribución normal estándar

PRECISIÓN DE LA ESTIMACIÓN POR IC. NIVEL DE CONFIANZA. RELACIÓN ENTRE AMBAS.

La expresión que se suma y se resta al estimador puntual para obtener el intervalo de confianza (d). Examinemos sus componentes:

$$d = Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

σ → desviacíon estándar de la población

$\sqrt{n}$ → Tamaño de la muestra

$Z_{1-\alpha/2}$ → Percentil del normal estándar también coeficiente de confianza" o "coeficiente de confiabilidad"

(1-α) Nivel de Confianza: probabilidad de que el verdadero valor del parámetro poblacional esté dentro del intervalo de confianza

La estimación será más precisa mientras más estrecho sea el intervalo, por tanto:
• A medida que aumenta el tamaño de la muestra (si los demás componentes se mantienen fijos) disminuye la amplitud del intervalo y la estimación es más precisa.
• A medida que el nivel de confianza es mayor aumenta la amplitud del intervalo y la estimación es menos precisa.

Ejemplo.

Vamos a suponer que en el ejemplo desarrollado se desconoce la varianza poblacional. A partir de los datos de la muestra se calcula s, que es la desviación estándar de la muestra y se obtuvo s= 9 latidos por minuto. Nuestro objetivo será el mismo, obtener el IC del 95% para la media poblacional de latidos por minuto, pero no podemos usar la fórmula:

$$\left\{\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

Límite inferior Límite superior

k	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617

Ahora se debe usar la fórmula del intervalo del (1- α)100% de confianza para la media poblacional cuando la varianza es desconocida

Límite inferior
Límite superior

$$\bar{X} - t_{k,1-\alpha/2}S/\sqrt{n}$$

$$\bar{X} + t_{k,1-\alpha/2}S/\sqrt{n}$$

Donde s es la desviación estándar de la muestra y tk,1-α/2 es el percentil que corresponda de la distribución t de student. La distribución t, al igual que la normal estándar, tiene su tabla

Li =

Ls =

Donde k = tamaño de la muestra menos 1

Calculando los valores de li y ls:

$\bar{x}$ = 90,

n=49 luego k= n-1= 48

$\sqrt{n}$ = 7 α=0,05 luego

1-α/2=0,975 s= 9

Entonces:

li = 90 – $t_{48;\,0,975}$ (9) /7
= 90 – (2,021) (9) /7
= 90 – 2,6
= 87,4

Ls= 90 + 2,6
= 92,6

Intervalo: [87,4;92,6]

Interpretación:
para un nivel de confianza del 95 %, la media poblacional de latidos del corazón fluctúa aproximadamente entre 87,4 y 92,6 latidos por minuto.

Para la proporción

Intervalo del (1-α)100% de confianza para la proporción poblacional P (aproximación a la normal)

$$ls = \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}\times\hat{q}}{n}} \qquad li = \hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}\times\hat{q}}{n}}$$

DONDE:

$\hat{p} = k/$ $\hat{q} = 1-\hat{p}$

k: número de elementos de la muestra con la característica de interés

Problema a solucionar mediante un intervalo de confianza.

Supongamos que se escogieron de manera aleatoria 93 mujeres embarazadas para realizar un estudio de salud bucal. Entre otros objetivos se quiere estimar la proporción poblacional de caries dentales en mujeres embarazadas a través de un IC del 95%. Se encontró que 24 de dichas mujeres tenían caries dentales, construya el intervalo.

Datos: k = 24, n = 93 luego $\hat{p} = 24/93 = 0,258$ $q = 1-0,258 = 0,742$

De la tabla: $Z_{0.975}$ = 1,96

$$li = 0,258 - 1,96\sqrt{\frac{0,258\times0,742}{93}} = 0,258 - 1,96\times0,045 = 0,169$$

ls = 0,258 + 1,96 X 0,045

El intervalo obtenido es [0,169; 0.347]
Puede escribirse también 0,169 ≤ p ≤ 0.347

Interpretación: La proporción poblacional de caries dentales en mujeres embarazadas fluctúa aproximadamente entre 0,169 y 0,347 con un nivel de confianza del 95%

==========

A diario el ser humano se enfrenta a la toma de decisiones.
El profesional médico también se enfrenta a la toma de decisiones, cuando tiene que discriminar entre alternativas de acción para decidir el tratamiento que se debe

seguir con un paciente, pero debe tener absoluta consciencia del riesgo que enfrenta si toma la decisión equivocada.
Por tanto, le es imprescindible tener elementos de base científica que le ayuden a escoger ante la incertidumbre la mejor alternativa

¿QUÉ ES UNA HIPÓTESIS ESTADÍSTICA?

- Una hipótesis estadística es una afirmación sobre el comportamiento de una variable aleatoria, susceptible de ser verificada.
- Es una suposición sobre los parámetros de la población:
  - Media
  - Proporción/porcentaje

La hipótesis estadística siempre estará integrada por la hipótesis nula y la alternativa, ambas forman una unidad.

- Hipótesis nula: H_0
- Hipótesis alternativa: H_1

Ejemplo: Un investigador quiere probar si el nivel medio de proteínas totales en los pacientes de cierta población es diferente de 68g/L.

H_0: m = 68g/l H_1: m (>, ≠, <) 68g/l

Hipótesis Nula. (H_0)

- Indica que un parámetro de una población (tal como la media, la proporción, etc.) es igual a un valor hipotético.
- Suele ser una afirmación inicial que se basa en análisis previos o en conocimientos especializados.
- Es un enunciado de que "no hay diferencias “por tanto siempre está presente el signo de igualdad.
- Basándose en los datos de una. muestra, la prueba determina si se puede o no rechazar la hipótesis nula por eso es la que contrastamos.

Hipótesis alternativa. (H_1)

- Es el enunciado que se desea poder concluir que es verdadero de acuerdo con la evidencia proporcionada por los datos de la muestra.
- En ella se considera lo que desea probar el investigador, pero teniendo en cuenta que el signo de igualdad siempre va en la hipótesis nula.

En la hipótesis nula (H_0) se afirma o considera la no ocurrencia del resultado esperado, mientras que la hipótesis alternativa (H_1) contradice, en algún sentido, la hipótesis nula. De acuerdo a esto, rechazar H_0 equivale a probar el resultado esperado.

PRUEBA HIPÓTESIS:

Es un método estadístico ideado con el objetivo de analizar si los datos obtenidos a través de una muestra observada de una variable aleatoria resultan o no contradictorios en relación con la hipótesis nula de la prueba para en base a esto tomar una decisión respecto al valor supuesto del parámetro en la hipótesis nula.
En general se trata de buscar un procedimiento para poder tomar una decisión cuando el problema al que nos enfrentamos es el de reconocer comportamientos diferentes entre poblaciones para alguna característica en particular

La decisión consiste en rechazar o no la hipótesis nula partiendo de la información de una muestra aleatoria. Hay que tomar una decisión en presencia de lo desconocido, pero hay ciertas regularidades que se deben cumplir:
• Los datos deben provenir de una muestra aleatoria simple.

• Se debe conocer la distribución de probabilidad que rige en la población original.
• Habrá que fijar a priori el error máximo que estamos dispuestos a aceptar cuando se tome la decisión.

Estadígrafo de prueba

- Es una variable aleatoria que se calcula a partir de los datos de una muestra aleatoria y se utiliza en una prueba de hipótesis para poder tomar la decisión de rechazar o no H_0.
- Mide el grado de concordancia entre la muestra y H_0.
- Resume la información de la muestra acerca del parámetro, que depende de un buen estimador que tenga una distribución muestral conocida.
- Su valor puede cambiar aleatoriamente según la muestra pero el valor particular observado depende de la única muestra estudiada.

FÓRMULA GENERAL:

$$\text{Estadígrafo de prueba} = \frac{\text{Estimador} - \text{Parámetro}}{\text{Error Estándar del Estimador}}$$

Tipos de errores

Al realizar una prueba de hipótesis estadísticas se pueden cometer dos tipos de errores:

- Error Tipo I: Se produce cuando la hipótesis nula es verdadera, pero se rechaza como consecuencia del contraste:
- Error Tipo II: Se comete cuando la hipótesis nula es falsa y, como consecuencia del contraste se acepta:

SE DECIDE	H_0 verdadera	H_0 falsa
No Rechazar	OK	ERROR TIPO II
Rechazar	ERROR TIPO I	OK

Nivel de significación α

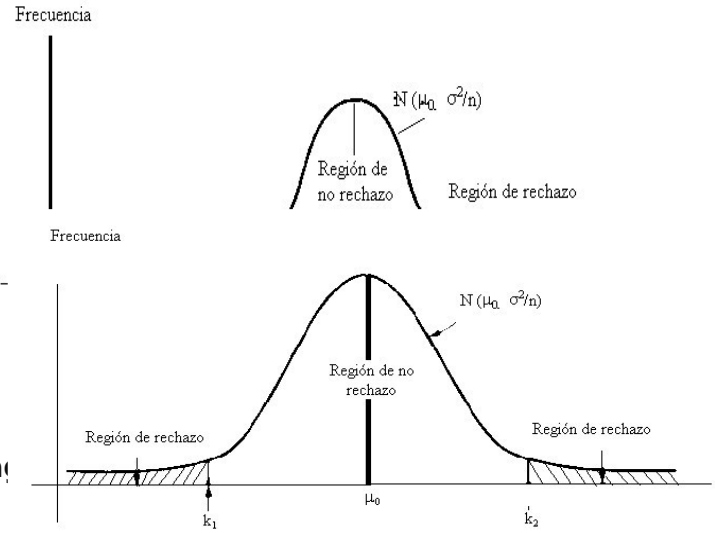
- Es el valor máximo de probabilidad que se está dispuesto a aceptar, para que ocurra el suceso de rechazar la hipótesis nula asumiendo que ésta es verdadera.
- Es el valor máximo de probabilidad que se está dispuesto a aceptar, para que ocurra el suceso de rechazar la hipótesis nula asumiendo que ésta es verdadera.

La costumbre es hacer **alfa=0.05 (5%) o alfa=0.01(1%)**

Región Critica

Es la región de la distribución muestral, donde el estadígrafo toma el conjunto de valores para los cuales se rechaza la hipótesis nula.

- El conjunto de valores muestrales que conducen a rechazar H0 suponiendo



que H0 es cierta se conoce como, región crítica o de rechazo, de la prueba de hipótesis.

- La región crítica o de rechazo de la PH estará en concordancia con la prueba que se esté aplicando, que puede ser unilateral o de una cola y bilateral o de dos colas.

Contrastes: unilateral y bilateral

La posición de la región crítica depende de la hipótesis alternativa
< unilateral izquierda > Unilateral derecha ≠ Bilateral

Sinónimos de la expresión "Estadísticamente significativo"

- Rechazo de la hipótesis nula
- Existe la suficiente evidencia para dudar de la hipótesis nula
- El resultado observado no es compatible con la hipótesis nula
- Es improbable obtener un resultado como el observado si la hipótesis nula es cierta
- Es improbable que el resultado observado sea debido al azar
- Las variaciones inherentes al muestreo no bastan para explicar el resultado observado
- Las muestras proceden de poblaciones diferentes

El test de significación estadística nos permite conocer la probabilidad de equivocarse si rechazamos la Ho, pero es evidente que la relevancia del fenómeno a estudiar es un
elemento clave en la toma de decisiones.

ALGORITMO GENERAL PARA REALIZAR UNA PRUEBA DE HIPÓTESIS

- Seleccionar los datos del problema, plantear las hipótesis, fijar el nivel de significación.
- Seleccionar y evaluar el estadígrafo de prueba adecuado.
- Determinar la región crítica y tomar la decisión de rechazar H_0 si el valor del estadígrafo cae en ella y no rechazarla en caso contrario.
- Se dice no rechazo la hipótesis nula, lugar de acepto la hipótesis nula, como argumento semántico, para no cometer el error de tipo II.
- Interpretar el resultado: decir si se corrobora lo que el investigador quiere probar

Ejemplo No. 1

Se quiere probar si el nivel medio de colesterol en una cierta población que se distribuye normal es diferente de 5,5mmol/l, se conoce la desv. estándar de la población que es de 2,85mmol/l. Para ello se extrajo una muestra de tamaño igual a 90 individuos y donde se halló la media muestral que fue de 5,25mmol/l.

1o. Seleccionar los datos del problema, plantear las hipótesis, y fijar el nivel de significación.

$\mu = 5.5$ mmol/l
$\sigma = 2.85$mmol/l
n = 90
$\bar{x} = 5.205$mmol/l

plantear las hipótesis:
$H_0: \mu = 5.5$ mmol/l
$H_1: \mu \neq 5.5$ mmol/l

es una prueba bilateral

Vamos a fijar un nivel de significación de un 5%

Prueba para la media μ de una población con distribución normal con varianza σ^2 conocida

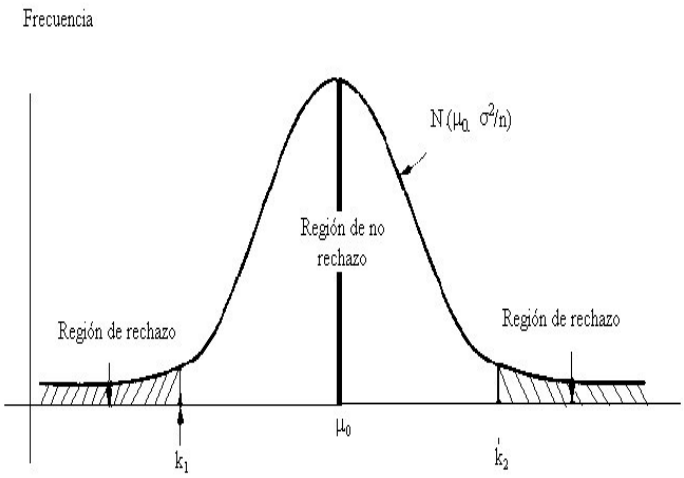
2o. Seleccionar y evaluar el estadígrafo adecuado Podemos calcular el valor: z que se conoce con el nombre de estadístico de la prueba de hipótesis. Al valor z también se le puede llamar observado". Noten que este valor es el resultado de estandarizar la media de la muestra suponiendo que lo que se cumple es la hipótesis nula.
Después de calculado, el z observado se compara con un valor Z teórico (percentil de z), que va a depender del nivel de significación de la prueba y al que también se le llama "valor crítico z".

$$Z=\frac{\overline{x}-\mu_0}{\sigma}\sqrt{n}=\frac{5{,}25-5.5}{2{,}85/\sqrt{90}}=-0{,}83$$

3º Determinar la región crítica y tomar la decisión de rechazar H_0 si el valor del estadígrafo cae en ella y no rechazarla en caso contrario.
En este caso los valores de Z critico son -1,96 y 1,96

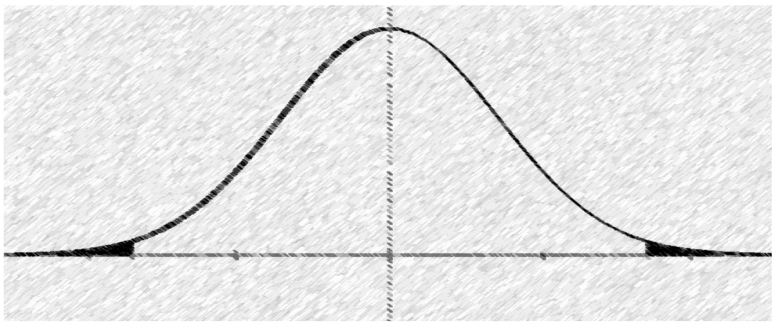
Determinar Región Crítica.
La región de rechazo se puede fijar antes de conocer el valor de la media muestral y de calcular Z por tanto se desconoce si Z va a caer hacia el extremo derecho o hacia el extremo izquierdo. En general esto se expresa diciendo que "se rechazará si está en los valores extremos con probabilidad 0.05 en cualquiera de las dos direcciones de la media poblacional

Hay que analizar si el estadígrafo calculado cae en la región crítica



Valores de Z crítico

Criterio de Decisión para este caso:
Si el módulo de z es mayor que el Z de la tabla (teórico), se rechaza H0 y se dice que z está en la región crítica o de rechazo de la prueba de hipótesis.



cae en la zona de no rechazo

-1.96 -0.83 1.96

No rechazo la hipótesis nula. El estadígrafo cae en la zona de no rechazo.

- Por tanto, no se puede corroborar que el nivel medio de colesterol de esa. Población sea diferente de 5,5 mmol/l porque no existen evidencias a favor de la hipótesis alternativa.

EJEMPLO 2:

Se conoce que en una población que se distribuye normal el nivel medio de hemoglobina es de 138 g/L, Existe la suposición de que en pacientes que padecen de un cierto tipo de anemia este valor es inferior. Para tratar de corroborar esta hipótesis se selecciona una muestra aleatoria de 9 pacientes que padecen de anemia y se obtiene un nivel medio de 130 g/L con una desviación estándar de 5,72 g/L ¿Se puede, en base a este resultado, afirmar que la suposición es acertada?

DATOS:
m = 138g/L
n = 9
s = 5.72 g/L

$\overline{x}=130\,g/L$

Planteando las Hipótesis

H_o $\mu = 138 g/L$

H_1 μ< 138 g/L

Es una prueba unilateral izquierda.

Vamos a fijar un nivel de significación de un 5% (0,05)

Es una prueba de hipótesis T está basada en el estadístico T, que sigue la distribución T student bajo la hipótesis nula.

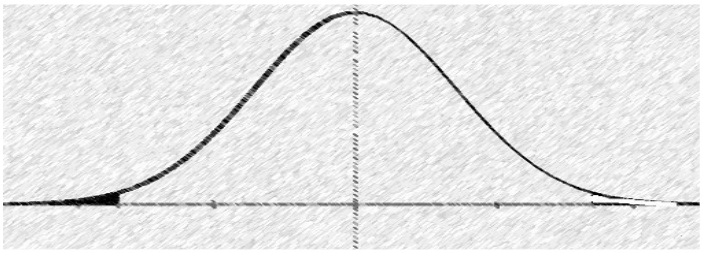
Podemos calcular el valor: T que se conoce con el nombre **de estadístico de la prueba de hipótesis**.
Al valor t calculado también se le puede llamar **T observado**". Esta prueba se utiliza cuando no
conocemos la desviación estándar de la población, pues la distribución T student solo depende del
tamaño de la muestra y no de la desviación estándar.
Después de calculado, el valor de T observado se compara con un valor T teórico o critico que lo
buscamos en la tabla evaluado en este caso para n-1; 1- Alfa

2o. Seleccionar y evaluar el estadígrafo adecuado

$$T=\frac{\overline{x}-\mu}{s}\sqrt{n}=\frac{130-138}{5,72\frac{}{\sqrt{9}}}=-4,19$$

3º Determinar la región crítica y tomar la decisión de rechazar H0si el valor del estadígrafo cae en ella y no rechazarla en caso contrario.

- En este caso el valor de t critico como es una prueba unilateral se evalúa para n-1; 1--Alfa es decir para 8; 0,995 siendo el valor T critico --3,355



Cae en la zona de rechazo
Rechazo la hipótesis nula. El estadígrafo cae en la zona de rechazo o región critica. Por tanto se puede corroborar la hipotesis alternativa de que el nivel medio de hemoglobina de los pacientes con ese tipo de anemia es un valor inferior.

PRUEBA PARA UNA PROPORCIÓN POBLACIONAL

Las enfermedades crónicas no transmisibles constituyen uno de los principales problemas de salud de la población cubana. Se sabe que el porcentaje de hipertensos cubanos mayores de 15 años es 19.5%. Un grupo de investigadores seleccionó una muestra aleatoria de 150residentes mayores de 15 años del municipio especial Isla de la Juventud para un estudio de morbilidad y encontraron un porcentaje de hipertensos de 18.2%. Diga si a partir de esta cifra puede afirmarse que en la Isla el porcentaje de hipertensos es menor que en el país. Utilice un nivel de significación de un 5%.

tendríamos que escribir la hipótesis estadística en función del párametro P, proporción poblacional de hipertensos 0,195

H_0: $P = P_0 = 0.195$ H_1: $P < P_0 = 0.195$

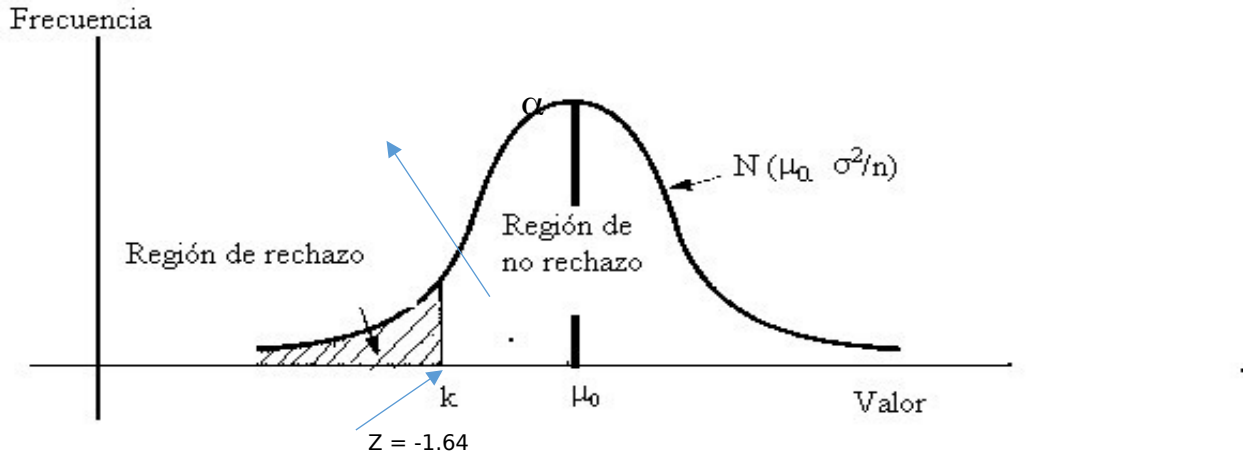
La proporción de hipertensos en la Isla (proporción muestral) es 0,182

Cuando tenemos proporciones se usa también Z observado que se halla a través de la

$$Z = \frac{p - P_0}{\sqrt{\frac{P0\,Q0}{n}}}$$

donde p es la proporción de hipertensos en la muestra y Q_0 el complemento de P_0

Región crítica para la prueba unilateral izquierda de una media poblacional, suponiendo que H0es cierta. Y para un Alfa de un 5%.



En el caso del ejemplo 2 tendríamos que escribir la hipótesis estadística en función del párametro P, proporción poblacional de hipertensos 0,195

H0: P = P 0 = 0.195 **H1: P ≠ P 0 = 0.195**

La proporción de hipertensos en la Isla (proporción muestral) es 0,182 Cuando tenemos proporciones se usa también Z que se halla a través de la fórmula

$$Z = \frac{p - P_0}{\sqrt{\frac{P0\,Q0}{n}}}$$

donde p es la proporción de hipertensos en la Muestra y Q0 el complemento de P0

De aquí que Z = (1,82 – 1,95)/$\sqrt{(0.809)(0.195)/150}$ = -0.013 /0,034 = -0,38

Ya sabemos que $Z_{0,975} = 1,96$ y como el módulo de z, 0,38 < 1,96 el estadístico está fuera de la región crítica por lo que con un nivel de significación del 0,05 no se rechaza la hipótesis de igualdad entre proporciones por lo que no puede afirmarse que en la Isla el porcentaje de hipertensos es menor que en el país.

Resumen: α, y criterio de rechazo

Sobre α

Es número pequeño, preelegido al diseñar el experimento
Conocido a sabemos todo sobre la región crítica
Podemos calcular el valor del estadígrafo y tomar la decisión.

Sobre el criterio de rechazo

analizar si el estadígrafo cae en la región crítica

Conclusiones

- Las hipótesisse plantean al inicio de la prueba realizar.
- En las pruebas de hipótesis se deben siempre plantear dos hipótesis la hipótesis nula y alternativa
- α debe de fijarse antes de realizar el experimento y debe ser pequeño
- El criterio de decisión para Rechazar una hipótesis consiste en observar si el estadígrafo de prueba cae en la región critica.
- Rechazar una hipótesis nula no prueba que sea falsa. Podemos cometer error de tipo I
- No rechazar una hipótesis no prueba que sea cierta. Podemos cometer error de tipo II
- Si decidimos rechazar una hipótesis debemos mostrar la probabilidad de equivocarnos.